Pipelines de ETL e Machine Learning com Apache Spark

Bibliografia, Referências e Links Úteis

Clustering
https://spark.apache.org/docs/latest/ml-clustering.html

Parquet Files
https://spark.apache.org/docs/latest/sql-data-sources-parquet.html